AUTOMATED DETECTION OF PIPE BURSTS AND OTHER EVENTS IN WATER DISTRIBUTION SYSTEMS

Abstract

This paper presents a new methodology for the automated near real-time detection of pipe bursts and other events which induce similar abnormal pressure/flow variations (e.g., unauthorised consumptions) at the District Metered Area (DMA) level. The new methodology makes synergistic use of several self-learning Artificial Intelligence (AI) techniques and statistical data analysis tools including wavelets for de-noising of the recorded pressure/flow signals, Artificial Neural Networks (ANNs) for the short-term forecasting of pressure/flow signal values, Statistical Process Control (SPC) techniques for short and long term analysis of the pipe burst/other event-induced pressure/flow variations, and Bayesian Inference Systems (BISs) for inferring the probability of a pipe burst/other event occurrence and raising corresponding detection alarms. The methodology presented here is tested and verified on a case study involving several UK DMAs with both real-life pipe burst/other events and engineered (i.e., simulated by opening fire hydrants) pipe burst events. The results obtained illustrate that it can successfully identify these events in a fast and reliable manner with a low false alarm rate.

Michele Romano, Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter, Devon, EX4 4QF, United Kingdom, Email: mr277@exeter.ac.uk

Zoran Kapelan, Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter, Devon, EX4 4QF, United Kingdom, Email: Z.Kapelan@exeter.ac.uk

Dragan A. Savić, Centre for Water Systems, College of Engineering, Mathematics and Physical Sciences, University of Exeter, Harrison Building, North Park Road, Exeter, Devon, EX4 4QF, United Kingdom, Email: D.Savic@exeter.ac.uk

INTRODUCTION

Water resources are under increasing pressure due to demographic, socioeconomic and environmental factors such as accelerated population growth, rapid urbanization, unsustainable consumption patterns, depletion and pollution of aquifers and more extreme environmental fluctuations due to the climatic consequences of global warming (Bates et al. 2008). In this scenario, the timely detection of pipe burst events in Water Distribution System (WDS) may enable the water companies to save large amounts of water and hence postpone or avoid the need to develop new resources in order to meet demand. The timely detection of pipe burst events in WDS also provides opportunities for water companies to save money, reduce their carbon footprint, achieve higher levels of operational efficiency and improve their customer service.

Despite all the past advances in pipe burst event detection methods since their introduction in the mid 1950s (Puust et al. 2010), there is still a need to further improve the efficiency and reliability of these methods. There is also a need to develop new and more effective burst event detection methodologies.

Equipment-based methods, such as leak noise correlators (Grunwell and Ratcliffe 1981) and pig-mounted acoustic sensing (Mergelas and Henrich 2005) are very effective in detecting pipe bursts and their technology has been improved considerably over the last few years. They are, however, expensive, labour-intensive, slow to run and may require the cessation of pipeline operations for long periods of time. On the other hand, transient-based methods (e.g., Kapelan et al. 2003; Misiunas et al. 2005) attempt to solve the detection problem by numerical modelling only. They, however, require a large number of measurement points, with high sampling frequency used to collect pressure and other measurements, which is very expensive. Also, these techniques often rely on complex and not so accurate transient network simulation models.

The latest developments in hydraulic sensor technology/on-line data acquisition systems have enabled water companies to deploy a larger number of more accurate and cheaper pressure and flow devices. As a result, large amounts of data are now being collected but not used much beyond for regulatory reporting. This is due to the volume and complexity of data received which exceeds human capability to analyse, interpret and extract useful information for operational and other purposes. To overcome this problem, a number of data analysis methods based on statistical and AI techniques have been developed recently (e.g., Mounce et al. 2010, 2011; Mounce and Boxall 2010; Aksela et al. 2009; Fenner and Ye 2011; Palau et al. 2011) with the aim of performing pipe burst event~~s~~ detection in WDSs by automatically processing pressure/flow data from the Supervisory Control And Data Acquisition (SCADA) system (i.e., signals).

Despite initial success, all these methods can be improved in terms of both pipe burst events detection reliability and detection time (e.g., the time elapsed between a pipe burst event occurrence and the generation of the corresponding alarm). This is achieved here by developing an automated detection system (hereafter referred to as the Event Recognition System – ERS) which is based on a new methodology that offers noticeable improvements over the aforementioned methods. The main improvements involve: (i) using advanced techniques for more efficient and effective processing of the hydraulic data gathered (e.g., wavelets for removing noise from the measured flow and especially pressure signals), (ii) taking advantage of a number of different ensembles of statistical and AI techniques each aimed at solving a different aspect of the pipe burst events detection problem (e.g., Statistical Process Control - SPC, and Artificial Neural Networks - ANNs for the analysis of the short/long term pressure/flow variations, and the recognition of the various types of evidence of an event occurrence), and (iii) using probabilistic inference engines (i.e., Bayesian Inference Systems - BISs) for reasoning under uncertainty, and for performing classification/inference by simultaneously (synergistically) analysing multiple event occurrence evidence and multiple pressure/flow signals at the DMA level. All these improvements result in more reliable and faster detection of pipe burst events in a WDS.

The paper is organised as follows. After this introduction, the relevant theoretical background and methodological detail of the data analyses performed by the automated ERS are reported in the methodology section. Then, the results obtained from applying the ERS to several DMAs in the United Kingdom (UK) are presented in the case study analyses section. Finally, the main conclusions are drawn and acknowledgements are given.

METHODOLOGY

Event Recognition System Overview

The methodology behind the automated ERS presented here aims at performing detection of pipe bursts and other events (e.g., illegal connections, sensor failures, unexpected water usages, etc.) at the DMA level and as they occur (i.e., in near real-time). Note that for the sake of simplicity, hereafter, "event" will be used as a generic term to indicate pipe burst events and other events which induce similar abnormal pressure/flow variations.

The current UK practice normally involves deploying pressure and flow sensors at the DMA entry/import/export points and a pressure sensor at the critical point in the DMA (i.e., the one located either at the point of highest elevation or alternatively at a location farthest away from the inlet). In the description that follows it is assumed that the DMA sensors collect the data at regular time intervals (e.g., 15 minute). These data are then communicated to the water companies at the same rate

as they are collected or at longer intervals (e.g., every 30 minutes), to improve sensor battery life. Also, data from different sensors is usually not synchronised in time.

The data processing in the ERS starts by receiving the data communicated by the DMA sensors. For each signal and at each communication interval *u* readings are obtained (e.g., 2 readings – assuming 15 minute sampled data, which are communicated every 30 minutes). These readings update a time series record which is stored in the Time Series database. Once the data from all the DMA pressure/flow sensors are fully processed as described below, the resulting *u* probability values that an event has occurred in the DMA and any additional information that may be used to perform a diagnosis of the incident occurring (e.g., to determine the likely cause of an alarm) are stored in the Alarms database. If any of the *u* probability values exceed a fixed detection threshold an alarm is generated.

Figure 1 to appear here.

The developed ERS enables event detection by performing the following three basic actions: (1) "capture" (i.e., learn/estimate) the expected patterns of the pressure/flow signals assuming that no event occurred in the DMA being studied - i.e. the Normal Operating Patterns (NOPs), (2) identify and estimate the event induced deviations between observed and "captured" DMA signal patterns, and (3) infer the probability that an event has occurred in the DMA based on the identified deviations.

Figure 1 shows a diagrammatic representation of the ERS. As it can be observed, the three basic actions described above (i.e., shown as dotted dashed rectangles) are performed in the ERS by making use of five subsystems (i.e., shown as solid snipped corner rectangles) each containing a number of different modules (i.e., shown as solid rectangles). The five ERS subsystems are as follows (1) the *Setup* subsystem, (2) the *Discrepancy Based Analysis (DBA)* subsystem, (3) the *Boundary Based Analysis (BBA)* subsystem, (4) the *Trend Based Analysis (TBA)* subsystem, and (5) the *Inference* subsystem.

The first ERS subsystem is used to perform the first basic action mentioned above (i.e., signal pattern capturing). The second, third and fourth ERS subsystems are used to perform the second basic action mentioned above (i.e., deviations identification/estimation). The reason for using three analysis subsystems is that, by making use of different ensembles of statistical/AI techniques, each of them focuses on recognising a specific type of evidence that an event has occurred. Furthermore, since they perform tasks in parallel they allow simultaneously assessing how an event affects the pressure/flow measurements from different perspectives (e.g., short-term and long-term effects). All this enables the three analysis subsystems to complement each other by: (a) providing more conclusive evidence of an event occurrence - different analysis subsystems may independently identify the deviations caused by

a particular event, and (b) enabling the identification of the deviations caused by different event types - each analysis subsystem looks at a particular deviation indicator that is more suitable than the others for identifying the deviations caused by a particular event type (e.g., the *DBA* and *BBA* subsystems are more suitable for sudden pipe burst events while the *TBA* subsystem is more suitable for gradually developing events such as pipe bursts developing gradually from background leaks). Finally, the fifth ERS subsystem is used to perform the third basic action mentioned above (i.e., event probability inference). Further details about the ERS subsystems and the associated modules can be found in the following sections.

Figure 1 also shows that the ERS has two main modes of operation, the "Assemble" mode and the "Execute" mode. The "Assemble" mode is used for 'tuning' the data-driven ERS when it is initialised (i.e., used for the first time in a DMA). Later on, it is used: (i) regularly (e.g., weekly) when the ERS is updated (to capture the latest normal operating conditions of a DMA) thereby providing a continuously adaptive ERS, and (ii) periodically when the ERS is reinitialised (following occasional operational/other DMA changes - e.g., re-valving). The "Execute" mode is the normal operating mode used at every communication interval to detect the events occurring and raise the alarms.

Setup Subsystem

Subsystem overview

The objective of the *Setup* subsystem is to perform initial data processing and "capture" different types of information about the NOP of the DMA signal being analysed. The data analyses in this subsystem are organised into five modules: (1) data retrieval module, (2) data pre-processing module, (3) statistics estimation module, (4) data de-noising module, and (5) ANN prediction model training & testing module. The first two modules are used for retrieving the historical signal data from the Time Series database and assembling a set of pressure/flow data that best represents the most recent NOP of the DMA signal being analysed (i.e., 'NOP data set'). The latter is then further processed in the third module to estimate several vectors of descriptive statistics (i.e., averages and standard deviations). These vectors provide a basic statistical information about the DMA signal NOP. The fourth module is used to remove the noise from the 'NOP data set'. At the end, the fifth module builds the ANN model for the short-term prediction of future DMA signal data. As this ANN prediction model assumes that no event occurred in the DMA being studied, it provides a model-based information about the DMA signal NOP. As can be observed from Figure 1, the *Setup* subsystem runs in the "Assemble" mode when the ERS is (re)initialised, and periodically updated (i.e., every l days - e.g., every 7 days).

Data retrieval module

This module is used for retrieving, for the DMA signal being analysed, of *m* days (e.g., 14 days) of past raw data (i.e., 'raw historical data set') from the Time Series database. The *m*-day 'raw historical data set' is then passed on to this subsystem's data pre-processing module (see Figure 1).

Data pre-processing module

This module is used for assembling the 'NOP data set'. This is achieved by using a two step procedure. The first step involves: (i) checking and correcting erroneous timestamps, (ii) creating a uniformly spaced time series, (iii) replacing blank entries with missing value indicators (NAN – 'not a number'), (iv) assigning Time of the Day (TofD) (i.e., a value between 1 and *g*, where 1 corresponds to midnight and *g* is the number of samples in one day) and Day of the Week (DofW) (i.e., a value between 1 and 7) values to each measurement, (v) rearranging the resulting *m*-day time series into *m* vectors (i.e., one vector for each day with *g* pressure/flow values), (vi) using a heuristics-based procedure to discard vectors containing large chunks of missing data, and (vii) using the linear interpolation to fill in the missing values in each of the remaining vectors (i.e., 'valid' days). As a result, an *n*-day 'repaired historical data set' is obtained. The second step involves gradually filtering out (from the 'repaired historical data set') outliers and measurements that are not consistent with the expected pressure/flow variations assuming that no event occurred in the DMA. This is achieved by using three statistical tests applied in sequence.

In the first statistical test, considering that each of the *n* vectors in the 'repaired historical data set' can be represented as (x_{1j}, x_{2j},…, x_{ij}) where $i = 1{:}\,g$ and $j = 1{:}\,n$, a vector of 'daily' averages ($\bar{x}_1$, $\bar{x}_2$,…, $\bar{x}_i$) and a vector of 'daily' standard deviations (σ_1, σ_2,…, σ_i) are computed. Then, each x_{ij} value is checked to see if it falls inside the interval ($\bar{x}_i - N_{l1^{st}}\ \sigma_i$, $\bar{x}_i + N_{u1^{st}}\ \sigma_i$), where $N_{l1^{st}}$ and $N_{u1^{st}}$ are user defined multipliers denoting the acceptable lower and upper confidence limits. If x_{ij} is outside these limits the entire *j* vector is discarded. As a result the number of 'valid' days may be reduced from *n* to *p*.

The second statistical test enables assessing how the mean of the remaining *p* 'valid' days changes over time. It involves the use of Control Charts (Shewhart 1931). A Control Chart is a graphical representation of descriptive statistics that can be used to track unusual variations in a process. A mean μ_k, where $k = 1{:}\,p$, is calculated for each of the remaining *p* 'valid' days of data. The Control Chart plots the *p* values μ_k in time order, a centre line at the average of the means $\bar{\mu}_k$, and upper and lower control limits at a user defined number $N_{l2^{nd}}$ and $N_{u2^{nd}}$ of standard deviations from the centre line. The standard deviation is estimated by taking the average of the *p* 'valid' days standard deviations. If μ_k is an outlier of the range defined by the upper and lower control limits, the k^{th} day is

defined as 'out of control' (Breyfogle 1999) and the k^{th} vector removed. After applying this second statistical test, the number of 'valid' days may be further reduced from p to q.

The third statistical test reapplies the first statistical test but this time considering the remaining q 'valid' days and new user defined multipliers for the acceptable lower and upper confidence limits $N_{l3^{rd}}$ and $N_{u3^{rd}}$, respectively. Once the third statistical test has been performed the ERS is potentially left with a dataset consisting of r 'valid' days (i.e., 'NOP data set').

Alternatively, if a 'repaired historical data set' containing a significant number of 'valid' days (i.e., $n \geq f$ - e.g., $n \geq 90$ days) can be assembled, the ERS groups the n 'valid' days according to their relevant DofW before performing the three statistical tests described above. The r-day 'NOP data set' is then put together by using the 'valid' days left in each DofW-group.

The reason for using the two alternative approaches described above is that, although the latter allows better accountability (from a statistical point of view) of the differences between the various DofWs (e.g., a weekday is likely to show a different pattern from a day during the weekend), its exclusive use would imply that the ERS has to remain inoperative for a significantly longer period of time when its (re)initialisation is required.

The assembled 'NOP data set' is then passed on to this subsystem's statistics estimation and data de-noising modules (see Figure 1).

Statistics estimation module

This module is used for "capturing" a basic statistical information about the NOP of the DMA signal being analysed. Depending on the number of 'valid' days n in the 'repaired historical data set': (i) one 'NOP average day' vector and one 'NOP daily standard deviations' vector are calculated (using the same procedure used in the first and third statistical tests described in the previous section) based on the r-day 'NOP data set' (i.e., $n < f$), or (ii) seven (i.e., one for each DofW-group) 'NOP average day' vectors and seven 'NOP daily standard deviations' vectors are calculated based on the 'valid' days left in each DofW-group after the three statistical tests have been performed (i.e., $n \geq f$).

The estimated 'NOP average day' and 'NOP daily standard deviations' vectors are then passed on to the evidence generation module of the *BBA* subsystem (see Figure 1) where they will be repeatedly used for establishing the boundaries within which the DMA signal data coming on-line from the relevant pressure/flow sensor should lie assuming that no event has occurred in the DMA.

Data de-noising module

This module removes noise from the assembled 'NOP data set' using the Wavelet Analysis (Meyer and Roques 1993). This is done because removing noise improves the prediction accuracy of the

ANN prediction model (Romano et al. 2010a). Also, de-noising of a pressure/flow signal can be performed without compromising its non-stationary or transitory characteristics, which are of particular importance in anomaly detection (especially when the de-noising of the data coming on-line is considered). Unlike Fourier analysis which consists of breaking up a signal into sine waves of various frequencies, Wavelet Analysis breaks up a signal into scaled (i.e., stretched/compressed) and shifted (i.e., delayed/hastened) versions of the original (or mother) wavelet. This dual frequency-time representation allows noise to be removed from signal frequencies that are likely to contain important information.

The conceptual details of the de-noising procedure applied here can be found in Donoho and Johnstone (1995). It involves the following steps: (1) perform a one level Discrete Wavelet Transform (DWT) of the original signal (Mallat 1989) resulting in two sub-signals: (i) approximation, which refers to the high-scale (i.e., low-frequency) component of the original signal, and (ii) detail, which refers to the low-scale (i.e., high-frequency) component of the original signal; (2) compute the Universal Threshold (Donoho and Johnstone 1995) and apply it to the detail sub-signal coefficients; and (3) perform Inverse Discrete Wavelet Transform (IDWT) using the original approximation sub-signal coefficients and the modified detail sub-signal coefficients.

The resulting 'de-noised NOP data set' is then passed on to the ANN training & testing module. Additionally, the Universal Threshold (computed and used in this module) is then passed on to the data de-noising module of the *DBA* subsystem where it will be repeatedly used for removing noise from the DMA signal data coming on-line (see Figure 1).

ANN training & testing module

This module is used for developing a historical model for the short-term prediction of the analysed pressure/flow signal values assuming that no event occurred in the DMA being studied, and for computing two key descriptive statistics to serve as a measure of the historical model prediction error's variability. An ANN model is chosen here for building a short-term pressure/flow prediction model because of the inherent complexity of the WDSs. The main aim is to exploit the ability of this powerful data modelling tool to model any function without explicit knowledge of the parameters involved. Furthermore, the ANNs efficiency in modelling and forecasting water consumption has been demonstrated in a large number of studies (e.g., Adamowski 2008).

A number of ANN structures, transfer functions, and training algorithms were tested for the problem at hand (not shown here). A feed-forward multilayer perceptron ANN (Bishop 1995), with a hyperbolic tangent transfer function for the neurons in the single hidden layer (i.e., hidden neurons) and a linear transfer function for the neuron in the output layer, trained by using the back-propagation method (Rumelhart et al. 1986) was eventually selected. The aforementioned tests were also used to

identify a suitable strategy for selecting the number of hidden neurons that allows striking a good balance between ANN learning and generalisation. The identified strategy involves selecting the number of hidden neurons by using the Neuroshell2 manual (1996) rule of thumb and using the Weight Decay Regularisation technique (Bishop 1995).

The ANN prediction model developed here has the following inputs: (1) a user defined number (i.e., *LagSize*) of past DMA signal values, and (2) the TofD and DofW indices associated with the forecasting horizon, converted into a field type (i.e., ones and zeros) form. The results obtained when analysing the ANN prediction models performance for different values of the *LagSize* are reported in Romano et al. (2010a). The output of the ANN prediction model is the predicted DMA signal value one-step ahead in time (e.g. next 15 minutes).

The ANN prediction model training set consists of a subset (i.e., *Train%* - e.g. 80%) of the 'de-noised NOP data set'. The remaining data (i.e., *Test%*) form the test set which is used to evaluate the performance of the ANN prediction model. The goodness-of-fit measure used here is the Nash-Sutcliffe index (Nash and Sutcliffe 1970). The closer the value of the Nash-Sutcliffe index to 1 the better.

In this module, the discrepancies between the training set's observed and predicted values are used to calculate the values of the mean μ_{ts} and of the standard deviation σ_{ts}. These values can be seen as measures of the ANN model prediction error's variability and are hence passed on to the evidence generation module of the *DBA* subsystem (see Figure 1). In that module, they will be repeatedly used for establishing the statistical limits within which the discrepancies between the DMA signal values coming on-line and their ANN model predicted counterparts should lie assuming that no event has occurred in the DMA being studied.

Additionally, the trained and tested ANN prediction model is passed on to the ANN forecasting module of the *DBA* subsystem (see Figure 1). In that module, it will be repeatedly used together with the DMA signal data coming on-line for performing one-step ahead predictions of the future DMA signal values.

Discrepancy Based Analysis Subsystem

Subsystem overview

The *DBA* subsystem is the first of the three ERS subsystems that aim at performing deviations identification and estimation. It provides a first way of doing this, which is based on the information about the NOP of the DMA signal being analysed "captured" by means of the ANN prediction model trained and tested in the *Setup* subsystem. Specifically, it checks that the discrepancies between the incoming observed DMA signal values and their ANN predicted counterparts do not exceed pre-

defined limits based on the estimated measures of the ANN model prediction error's variability (i.e., μ_{ts} and σ_{ts}). This way, it focuses on the identification of the pressure/flow deviations induced by sudden events and it can identify small to large deviations. As a matter of fact, the size of the minimum identifiable deviation is theoretically of the same order of magnitude as σ_{ts}. Because of all this, the *DBA* subsystem is particularly well suited for identifying the beginning and end of an event. The data analyses in this subsystem are organised into four modules: (1) data retrieval module, (2) data de-noising module, (3) ANN forecasting module, and (4) evidence generation module. The first two modules are used for retrieving from the Time Series database and de-noising the DMA signal data coming on-line. Then, the third module is used for performing one-step ahead prediction of the DMA signal values. Finally, the fourth module is used for: (i) comparing the incoming observed DMA signal values to the signal values predicted by the ANN model, (ii) identifying/estimating significant (i.e., indicative of an event occurrence) discrepancies between those values, and (iii) further processing the identified discrepancies to provide reliable evidence of an event occurrence. As can be observed from Figure 1, this subsystem runs in the "Execute" mode every *a* minutes (e.g., every 30 minutes – depending on the data communication frequency).

Data retrieval module

This module is used for retrieving (from the Time Series database) the latest *b* raw pressure/flow values for each DMA signal being analysed. Once this is done, each of these values is assigned its TofD and DofW values. Note that *b* is the number of past pressure/flow values that enables: (1) the discrepancies between the incoming observed DMA signal values and their ANN predicted counterparts to be analysed over consecutive time steps (see this subsystem's evidence generation module), and (2) the resulting pieces of event occurrence evidence (which are relative to multiple successive time steps) to be synergistically processed by the Signal level BIS and DMA level BIS modules in the *Inference* subsystem. Bearing this in mind, the value of *b* is determined automatically by the ERS by taking into consideration the following: (i) the *LagSize* used for training and testing the ANN prediction model, (ii) the maximum number of consecutive time steps analysed in this subsystem's evidence generation module, and (iii) the number of time steps considered by the Signal level BIS and DMA level BIS modules in the *Inference* subsystem. For example, given that a maximum of 8 consecutive time steps are analysed in this subsystem's evidence generation module (see below) and assuming that the *LagSize* is equal to 4 and that 3 time steps are considered in the Signal level BIS and DMA level BIS modules, the value of *b* is equal to 15 (= 4 + 8 + 3). This set of values forms the 'raw incoming data set', which is then passed on to this subsystem's data de-noising module (see Figure 1).

Data de-noising module

This module is used for removing noise from the 'raw incoming data set' while preserving and isolating the characteristics of an event occurring. This is achieved by using the same procedure and de-noising threshold described/computed in the data de-noising module of the *Setup* subsystem. The reason for using that de-noising threshold is that it is based on the 'NOP data set', therefore it is "tailored" to the noise level assuming that no event occurred in the DMA being studied. The resulting 'de-noised incoming data set' is then passed on to the ANN forecasting module (see Figure 1).

ANN forecasting module

This module is used for obtaining the predicted DMA signal values corresponding to the latest *b-LagSize* values in the 'de-noised incoming data set'. This is achieved by using the pressure/flow values in the 'de-noised incoming data set' as input to the ANN prediction model developed in the ANN training & testing module. The resulting 'predictions data set' is then passed on to this subsystem's evidence generation module (see Figure 1).

Evidence generation module

The objectives of this module are: (a) to reliably identify the occurrence of event-induced deviations from the NOP of the analysed DMA signal, and (b) to generate, at each relevant time step, an evidence that reflects the event-induced deviation presence (or absence) and its severity (if present). These objectives are achieved by: (i) analysing the discrepancies between the DMA signal values in the 'predictions data set' and their observed counterparts in successive time steps using SPC-based Control Rules (Shewhart 1931), and (ii) determining and classifying, at each relevant time step, the number of Control Rule violations. The resulting vector of evidence is then forwarded as an input into the relevant BIS modules (i.e., Signal and DMA level) of the *Inference* subsystem (see Figure 1).

A simple Control Rule can be used to test if a discrepancy $x_{obs,t} - x_t$ falls outside a confidence interval $(\mu_{ts} - N_l\sigma_{ts}, \mu_{ts} + N_u\sigma_{ts})$ where $x_{obs,t}$ is a pressure/flow measurement at time t, x_t is the corresponding value predicted by the ANN model, μ_{ts} and σ_{ts} are the mean and the standard deviation of the discrepancies between the training set's observed and predicted values, and N_l and N_u are user defined multipliers denoting the acceptable lower and upper confidence limits. If, at any time step, a discrepancy falls outside these limits this indicates the presence of an event-induced deviation from the NOP of the DMA signal being analysed. However, if the same observation is made in consecutive time steps, it increases the chance of an event actually occurring.

This said, to further increase the effectiveness and reliability of the deviations identification, the following modified subset of the Western Electric Control Rules (Western Electric Company 1958) for detecting 'out of control' situations is used here: (i) if any discrepancy falls outside the $\pm 4\sigma_{ts}$ limits, (ii) if two out of three consecutive discrepancies fall above the $+3\sigma_{ts}$ limit or below the $-3\sigma_{ts}$

limit, (iii) if four out of five consecutive discrepancies fall above the $+2\sigma_{ts}$ limit or below the $-2\sigma_{ts}$ limit, and (iv) if eight consecutive discrepancies fall above the $+1\sigma_{ts}$ limit or below the $-1\sigma_{ts}$ limit. The modified set of Control Rules was identified after relevant sensitivity analysis (Romano et al. 2010b). The same analysis also tested different types of comparative variables and concluded that the discrepancy shown above generated the best event detection results.

The above Control Rules are applied to each discrepancy and a number of Control Rule violations is determined. This number is then classified as follows: (i) high (3 or 4 violations), (ii) moderate (1 or 2 violations), or (iii) none (0 violations).

Boundary Based Analysis Subsystem

Subsystem overview

The *BBA* subsystem is the second of the three ERS subsystems that aim at performing deviations identification and estimation. It provides another, second way of doing this based on the basic statistical information about the DMA signal NOP "captured" in the *Setup* subsystem. Specifically, it checks that the incoming observed signal values lie inside a "data envelope" whose boundaries are defined by using the vectors of descriptive statistics estimated from the 'NOP data set' (i.e., 'NOP average day' and 'NOP daily standard deviations' vectors). Similarly to the *DBA* subsystem, this subsystem focuses on the identification of the pressure/flow deviations induced by sudden events. However, it can only identify medium to large deviations. Indeed, at each time step (i.e., a particular TofD), the magnitude of the minimum identifiable deviation is theoretically of the same order of magnitude as three times the corresponding standard deviation value in the relevant 'NOP daily standard deviations' vector (see this subsystem's evidence generation module). Despite this, it complements the *DBA* subsystem by providing further evidence of an event occurrence. This is because, by analysing the incoming observed DMA signal values as described above, this subsystem has the potential to enable identifying deviations throughout the duration of an event. The data analyses in this subsystem are organised into two modules: (1) data retrieval module, and (2) evidence generation module. The first module is used for retrieving the incoming DMA signal data from the Time Series database. Then, the second module is used for: (i) checking whether the incoming observed DMA signal values lie inside the aforementioned statistical boundaries, (ii) identifying/estimating the event-induced excursions outside those boundaries, and (iii) further processing the identified excursions to provide reliable evidence of an event occurrence. As it can be observed from Figure 1, this subsystem also runs in the "Execute" mode every *a* minutes.

Data retrieval module

This module is used for retrieving (from the Time Series database) the latest *c* raw pressure/flow values of the DMA signal being analysed and associating to each measured value its TofD and DofW.

For the same reason given in the section describing the data retrieval module of the *DBA* subsystem with regard to the value of *b*, the value of *c* is determined automatically by the ERS. However, it only depends on: (i) the maximum number of consecutive time steps analysed in this subsystem's evidence generation module, and (ii) the number of time steps considered by the Signal level BIS and DMA level BIS modules in the *Inference* subsystem. This set of values forms the 'incoming data set' which is then passed on to this subsystem's evidence generation module (see Figure 1).

Evidence generation module

The objectives of this module are identical to the objectives of the evidence generation module of the *DBA* subsystem. However, the way these are achieved is as follows.

Firstly, data analyses are performed to check whether each DMA signal value in the 'incoming data set' exceeds the statistical boundaries. As each DMA signal value has an associated TofD and DofW, these boundaries are computed by using the corresponding values of the average, μ_i, and of the standard deviation, σ_i, in the relevant 'NOP average day' and 'NOP daily standard deviations' vectors. Specifically, the boundaries are defined as $\mu_i \pm N\sigma_i$, where N is a user defined multiplier. Once this is done, the identified excursions outside the statistical boundaries are identified/estimated and analysed in successive time steps using SPC-based Control Rules. Next, the number of Control Rule violations is classified into three categories and the resulting vector of evidence forwarded as an input into the relevant BIS modules of the *Inference* subsystem (see Figure 1).

Note however that, based on the results obtained after relevant sensitivity analysis (not shown here), the following set of Control Rules is used in this module: (i) any single observed value falls above the $+3\sigma_i$ boundary or below the $-3\sigma_i$ boundary, (ii) two out of three consecutive observed values fall above the $+3\sigma_i$ boundary or below the $-3\sigma_i$ boundary, (iii) four out of five consecutive observed values fall above the $+3\sigma_i$ boundary or below the $-3\sigma_i$ boundary, and (iv) eight consecutive observed values fall above the $+3\sigma_i$ boundary or below the $-3\sigma_i$ boundary.

Trend Based Analysis Subsystem

Subsystem overview

The *TBA* subsystem is the last of the three ERS subsystems that aim at performing deviations identification and estimation. It provides a further way of doing so which is neither dependent on the model-based nor the basic statistical information about the DMA signal NOP "captured" in the *Setup* subsystem. This subsystem focuses on the identification of the pressure/flow deviations induced by gradually developing events (e.g., pipe bursts developing gradually from background leaks, cumulative effect of background leaks, etc.). However, it also supports the other ERS analysis subsystems in the identification of the pressure/flow deviations induced by sudden events. The data

analyses performed in this subsystem are organised into three modules: (1) data retrieval module, (2) data pre-processing module, and (3) evidence generation module. The first two modules are used for: (i) retrieving from the Time Series database the latest historical signal data measured during a particular time window during the day, (ii) dealing with the erroneous timestamps and the missing data, and (iii) filtering out those time windows containing outliers and/or measurements that are not consistent with the expected DMA signal values assuming that no event occurred in the DMA being studied. At the end, the third module is used for: (i) monitoring on a Control Chart the trend of the average of the measurements in each of the remaining time windows, and (ii) identifying 'out of control' situations indicative of the longer term event-induced pressure/flow variations. As it can be observed from Figure 1, this subsystem runs in the "Execute" mode every *d* hours (e.g., 4 hours).

Data retrieval module

This module is used for retrieving from the Time Series database the latest *e* days (e.g., 14 days) of raw past DMA signal data measured during the *d*-hour (e.g., 4 hours) window to be analysed (e.g., latest 14 days of pressure/flow measurements recorded between midnight and 4 am, 4 am and 8 am, etc.). The retrieved data form the 'raw historical *d*-hour window data set' which is then passed on to this subsystem's data pre-processing module (see Figure 1).

Data pre-processing module

This module is used for assembling (starting from the 'raw historical *d*-hour window data set') a set of data that, assuming that no event has occurred in the DMA being studied, best represents how the pressure/flow measurements recorded during a *d*-hour window vary over time (i.e., 'NOP *d*-hour window data set'). This is achieved by using a data pre-processing procedure similar to that used in the data pre-processing module of the *Setup* subsystem. The main differences are as follows: (i) after correcting the erroneous timestamps and replacing the blank entries with missing value indicators, the data are rearranged to form *e* vectors (i.e., one for each *d*-hour window) each containing $d \times v_{ph}$ pressure/flow values, where v_{ph} (i.e., values per hour) is the number of pressure/flow values recorded in one hour (e.g., $v_{ph} = 4$ - if a 15 minute sampling rate is considered), and (ii) only one statistical test (similar to the first and third statistical tests described in the data pre-processing module of the *Setup* subsystem) is performed by using $N_{l4^{th}}$ and $N_{u4^{th}}$ as user defined multipliers. The rationale for using only one statistical test is allowing more variability in the assembled data set. The resulting 'NOP *d*-hour window data set' is then passed on to this subsystem's evidence generation module (see Figure 1).

Note however that the above procedure for dealing with the missing values and the statistical test are applied to all but the e^{th} vector (i.e., the most recent *d*-hour window). This is because the focus of the

analysis performed in the following evidence generation module is to identify the event-induced pressure/flow variations that may have been recorded during this particular time window.

Evidence generation module

The objectives of this module are similar to the objectives of the evidence generation modules of the other analysis subsystems. The main differences are as follows: (a) the evidence is generated every *d* hours, and (b) the evidence only reflects the presence (or absence) of an event-induced pressure/flow variation. These objectives are achieved by: (i) using a Control Chart to monitor how the mean of the pressure/flow measurements relative to the particular *d*-hour window being analysed changes over time, and (ii) determining and classifying the occurrence of an 'out of control' situation. Similarly to the other evidence generation modules, the evidence is then forwarded as an input to the relevant BIS modules of the *Inference* subsystem (see Figure 1).

In the analysis carried out here, a mean is calculated for each of the remaining *d*-hour vectors in the 'NOP *d*-hour window data set'. The Control Chart plots these means in time order, a centre line at the average of the means, and upper and lower control limits at three standard deviations from the centre line. The standard deviation is estimated by taking the average of the remaining *d*-hour vectors' standard deviations. If the mean of the measurements recorded during the latest *d*-hour time window is an outlier of the range defined by the upper and lower control limits, those pressure/flow measurements are considered as indicators of an 'out of control' situation. This information is then classified as either 'out of control' or 'in control'.

Inference subsystem

Subsystem overview

The *Inference* subsystem constitutes the final processing stage of the ERS as it enables performing the third basic action mentioned in the overview section (i.e., event probability inference). The data analyses performed in this subsystem are organised into two modules: (1) Signal level Bayesian Inference System module, and (2) DMA level Bayesian Inference System module. In these two modules all the generated evidence of an event occurrence at hand are combined to infer the probability that an event has occurred in the DMA, raise detection alarms, and provide additional information that may allow performing diagnosis of the event occurring. As it can be observed from Figure 1, this subsystem runs in the "Execute" mode every *a* minutes.

Signal level Bayesian Inference System module

The objective of this module are: (a) to calculate, at each time step during the data communication interval, the probability that an event has occurred (i.e., Signal level event occurrence probability) based on the evidence resulting from the three ERS analysis subsystems (see Figure 1), and (b) to

provide, at each time step during the data communication interval, a measure of the pressure/flow deviation from the NOP of the DMA signal being analysed "captured" by means of the basic descriptive statistics in the *Setup* subsystem. These objectives are achieved by: (i) using a Bayesian Network (BN) (Edwards 2000) for performing inference, and (ii) estimating the difference between an observed DMA signal value and its corresponding value in the relevant 'NOP average day' vector. Note that the latter measure of deviation may be used (after a detection alarm is raised – see the DMA level BIS section) to evaluate the likely event magnitude (e.g., burst flow). The module's output is then stored in the Alarms database.

A BN is used for achieving this module's first objective because it allows reasoning under uncertainty. A BN is ideally suited for situations that require combining multiple inputs in order to infer a meaningful (probabilistic) output for decision-making. Furthermore, it allows updating the probability of an outcome (e.g., event occurrence) as evidence accumulates. With enough evidence, it should become very high or very low. This said, the structure of the Signal level BIS is designed here in such a way that it is able to not only process synergistically the evidence from the different ERS analysis subsystems (i.e., *DBA/BBA/TBA*) but also to encode the temporal sequence of the incoming data. The latter is achieved by considering the event occurrence evidence relative to a number of time steps, z, which, in turn, allows reasoning over time by updating the probability of an event occurrence over consecutive time steps.

Although this module is not used to raise the detection alarms, it provides information that may be used in the diagnosis of an event occurring. In fact, it allows the user to assess the extent to which the particular signal has contributed to the generated alarm. The latter is particularly useful because it may allow the user of the ERS: (1) to identify the area of the network that has been affected the most by the event, and (2) to determine the likely cause of the event (e.g., faulty sensor or failure of a pressure modulating valve).

DMA level Bayesian Inference System module

The main objective of this module is to raise detection alarms at the DMA level if and when necessary. Similarly to the Signal level BIS, a BN which encodes the temporal sequence of the incoming data is used for inferring, at each time step during the data communication interval, the probability that an event has occurred (i.e., DMA level event occurrence probability). The main difference is that the DMA level BIS takes as an input the evidence resulting from the different ERS analysis subsystems coming simultaneously from all the DMA signals (see Figure 1). It has to be stressed that, by enabling the processing of all the event occurrence evidence available in a synergistic way it is expected that the ERS will be able to increase its efficiency and reliability of detecting the

occurrence of an event in the DMA being studied. Finally, a detection alarm is generated when the DMA level event occurrence probability goes above a user defined detection threshold (λ).

In addition to the above DMA level event occurrence probability, the DMA level BIS also calculates the event occurrence probabilities based on: (i) the evidence from the *DBA* subsystems only, (ii) the evidence from the *BBA* subsystems only, and (iii) the evidence from the *TBA* subsystems only. These three additional event occurrence probabilities can help evaluate the extent to which a particular type of analysis has contributed to generate an alarm and, for example, identify the event as a growing rather than a sudden pipe burst.

Once a detection alarm is raised, it is stored in the Alarms database and the following information is saved: (i) the DMA location, (ii) the DMA level event occurrence probability, (iii) the event occurrence probabilities for each type of analysis performed (i.e., short-term *DBA* and *BBA*, and long-term *TBA*), and (iv) the detection alarm start time. Note that, in order to avoid raising unnecessary alarms for the same event in the future, the ERS simply suppresses any further detection alarm for the user specified *'alarm inactivity time'* (e.g., 1 week).

CASE STUDY ANALYSES

Description

The methodology behind the ERS presented was tested and verified on: (i) Engineered Events (EEs) where fire hydrants were opened to simulate different types of pipe burst, and (ii) real-life pipe burst/other events.

In the first case, the data analyses performed made use of the pressure and flow data recorded by the sensors deployed in a UK DMA during July-August 2009 and March 2010 (as the EEs were conducted in these 3 months). This DMA is predominantly urban, has a total of 2,640 domestic properties and 500 commercial users. The total mains length is 24 km. It is equipped with a pressure and flow sensor at the DMA inlet as well as one pressure sensor at the critical DMA point.

In the second case, the data analyses performed made use of the pressure and flow data recorded by the sensors deployed in five UK DMAs during the eleven month period from July 2009 to June 2010. The selected DMAs have different characteristics and varying sizes. As an ensemble, they contain light industrial, urban, semi rural, and rural regions. The number of domestic properties in each of them varies between 409 and 3,493. The number of commercial users varies between 11 and 231. Their individual total mains length varies between 6.3 and 30 km. These DMAs are equipped with a flow sensor at the DMA inlet and a pressure sensor at the critical point in the DMA and, in one case only, with a flow and pressure sensor at the DMA inlet only.

In both cases, the DMA sensors recorded data at 15 minute intervals. The flow data were averaged values during the 15 minute sampling interval whilst the pressure data were 15 minute instantaneous values.

Note that although historical data were used for the data analyses performed here, the pressure and flow measurements were fed to the ERS in a simulated 'on-line' fashion (i.e., as the ERS would have been used in real-life). Table 1 summarises the user defined ERS parameters used for the case study analyses.

Table 1. to appear here.

Engineered Burst Events

In the analyses carried out here, the methodology behind the ERS was tested and verified on the EEs where pipe burst events were simulated by opening fire hydrants in the DMA being studied. This way, all "burst" related information (location, timing and approximate flow rate) was known which, in turn, enabled the evaluation of the methodology's performance based on comparisons between the ERS detection times and the corresponding actual hydrant opening times. In addition to this, the Receiver Operating Characteristics (ROC) graph (Egan 1975) and related metrics, i.e., area under the ROC curve – AUC (Hanley and McNeil 1982) were used too. Note that ROC graphs have long been used in signal detection theory to depict the trade-off between hit rates and false alarm rates of classifiers. The AUC of a classifier represents the probability that the classifier will score a randomly drawn positive sample higher than a randomly drawn negative sample. The closer the AUC value to 1 the better.

A total of 9 EEs were carried out as follows: 3 events in July 2009, 3 events in August 2009, and 3 events in March 2010. In order to simulate different types of burst, hydrants at different location were opened achieving different flow rates. Note that any single EE lasted one day maximum and that different EEs were sometimes carried out during the same week. Having said this, the *'alarm inactivity time'* parameter had to be chosen carefully. Choosing it as one week, for example, would have implied that only one alarm could have been raised for multiple EEs carried out during the same week. Thus, for the purposes of the analyses that evaluated the capabilities of the event detection methodology based on comparisons between the detection times and the corresponding actual hydrant opening times, it was set as equal to 1 day.

Table 2 shows the detection times obtained and the corresponding actual hydrant opening and closing times for all the EEs. The actual flow rates and their corresponding values estimated by the ERS are also given together with the values of the actual flow rate expressed as percentage of the average DMA inflow. It can be observed that all EEs were successfully detected. The underlined alarm start times refer to those events that were detected at the best possible time (within the 15-minute sampling

rate). Alarm start times in normal font refer to those events that were detected with a delay not greater than one hour, whilst alarm start times in bold refer to those events that were detected with delays longer than 1 hour. The main reason for the comparatively long delay in detection for the EE carried out on the 1st of March 2010 is that the DMA pressure signals were only mildly affected by this event (most likely due to the location of the hydrant opened being far away from the pressure sensors). As a result, until late evening that day (when the "legitimate" variability in pressure/flow became less accentuated), the pieces of event-occurrence evidence generated by the three ERS analysis subsystems were not sufficient to produce a DMA level event occurrence probability greater than the user defined detection threshold (i.e., greater than 0.5) and hence trigger an alarm. With regard to the comparatively long delay in detection for the EE carried out on the 2nd of March 2010, on the other hand, note that only 10 minutes separate this EE from the EE carried out on the 1st of March 2010. Therefore, with the relevant value for the '*alarm inactivity time*' parameter, it was not possible to correctly identify its start time although the ERS generated a DMA level event occurrence probability of 0.68 at 08:00 that day (i.e., detection at the best possible time). The results obtained confirm that the event detection methodology presented here can effectively detect burst events. Furthermore they show that timely detections are achieved (i.e., within one hour for all but one of the simulated burst events).

Table 2. to appear here.

As an example, Figure 2 shows the results obtained by the ERS when the first engineered event is considered. Note that only the trend of the flow signal is shown in the figure. In this figure the vertical axis on the right shows the DMA level event occurrence probability (i.e., a value between 0 and 1). The leftmost vertical line indicates the time the detection alarm was raised and the vertical light bars (i.e., P global) indicate the DMA level event occurrence probability (i.e., only if greater than the user defined detection threshold, λ - i.e., greater than 0.5) at every time step (i.e., every 15 minutes). The rectangle in the top of part of the figure represents the duration of the EE.

Figure 2. to appear here.

For the purposes of the analyses that evaluated the capabilities of the event detection methodology using the ROC graph and the AUC value, the value of the '*alarm inactivity time*' parameter was set as equal to 15 minutes (i.e., all the DMA level event occurrence probabilities greater than λ raised detection alarms). This is because true and false alarm rates are obtained by comparison, at every time step, between the status of the hydrant (i.e., hydrant opened/hydrant closed) and the output of the DMA level BIS (i.e., DMA level event occurrence probability greater/smaller than λ). Results obtained for the three months studied were considered as a single set of data. Then, because the exact

hydrants' opening and closing times were known, a target vector consisting of a binary value (i.e., hydrant opened/hydrant closed) for each time step was associated to it.

Figure 3 shows the ROC curve for the three month data set. The figure also shows the detection threshold values printed at the corresponding curve positions. The resulting ROC curve is a two-dimensional depiction of the DMA level BIS performance (i.e., trade-off between true and false alarm rates). However, it is possible to reduce ROC performance to a single scalar value representing expected performance by calculating the area under the ROC curve. In this case the calculated AUC value is 0.88. The resulting ROC graph and its AUC value provide strong evidence of the good classification performance of the DMA level BIS.

By observing Figure 3, a few considerations about the choice of the best detection threshold value can be made. On the one hand, choosing a low value for the detection threshold (e.g., 0.4) has the potential to improve the methodology's sensitivity to the events. As an example, when the first EE carried out in March is considered (i.e., the one detected with an eleven hour and forty five minutes delay – see Table 2), a threshold value equal to 0.4 would have led to the detection of the simulated burst event at the best possible time (not shown here). However, this choice would have led to raising false alarms also. On the other hand, choosing a high value for the detection threshold (e.g., 0.6) produces a false alarm rate close to zero but may have allowed detecting fewer events only. Based on these observations, it is possible to state that a detection threshold value of 0.5 allows achieving good detection performance (see also the results of the tests on the real-life events). However, it does not represent the optimal choice. Therefore, the selection of the most suitable detection threshold value has to be ultimately based on the operator's decision and remains subject to the operator's confidence and expertise. In this scenario, the data analyses described above can be used to help the operator to make more informed decisions.

Figure 3. to appear here.

Real-Life Events

When the ERS was tested on real-life events the value of the '*alarm inactivity time'* parameter was set as equal to 1 week. With this setting, the ERS raised a total of 38 detection alarms.

In order to evaluate the performance of the developed methodology, the following data were available: (i) Main Repair (MR) records from the Work Management System (WMS) database (i.e., containing records of repairs carried out in the network), and (ii) Customer Contacts (CCs) records (i.e., customer complaints about potential problems related to water supply). The 17 MR records in the analysed eleven month period were considered as an indication of the occurred burst events. However, they only gave information about the creation date of the MR records not when the actual

MR jobs were completed. Furthermore, due to the nature of underground pipe bursts, the exact start date and time of these events could not be determined based on this information only. Therefore, a careful visual inspection of the data (supported by the MR and CC records) was carried out in order to determine their likely duration. This visual inspection of the data also allowed the identification of 7 clusters of CCs, which were not accompanied by any MR record. They were considered as an indication of pressure/flow anomalies whose exact cause is uncertain (e.g., illegal water usage, unusual system activity, operational DMA changes, etc.), but for which a record exist. Furthermore, it allowed the identification of 1 sensor failure event (i.e., characterised by the distinctive trend signature), and 5 other visible pressure/flow anomalies which were, however, not accompanied by any CC or MR records (i.e., did not impact the customers). All these events formed the set of real-life incidents against which the 38 detection alarms raised by the ERS were compared in order to see if a correlation existed (i.e., genuine alarms) or not (i.e., false alarms). Figure 4 shows the results obtained from this comparison.

Figure 4. to appear here.

It is possible to observe from this figure that 21 alarms were correlated to the 17 burst events, 6 alarms were correlated to the pressure/flow anomalies that generated several CCs, 3 alarms were correlated to the sensor failure event, 5 alarms were correlated to the 5 visible pressure/flow anomalies that did not affect the customers, and 3 were false alarms. Note that because of the *'alarm inactivity time'* parameter used, multiple detection alarms were sometime related to the same events. Furthermore, some event caused alarms in different DMAs. These results show that the event detection methodology was able to detect all but one of the 30 real-life events identified by means of the visual inspection of the data. Furthermore, it did so with a low false alarm rate (i.e., about 8%). It is therefore possible to state that these early results point to the potential of the methodology presented here to effectively and reliably detect pipe bursts and other real-life events.

It is worth noting that the alarm rates (false alarms in particular) and the detection times obtained by using the methodology presented here on both engineered pipe burst events and real-life pipe burst/other events should be, ideally, compared with corresponding results obtained by using other methodologies available. For example, Mounce et al. (2010) and Mounce and Boxall (2010) reported detection results obtained from the near-real time analysis of flow data in 144 DMAs for a 2-month trial period and 146 DMAs for a 1-year trial period, respectively. The detection times reported form these two case studies are of the order of 12 hours and false alarms rates are 15% and 18% respectively. In another study, Mounce et al. (2011) reported faster alarm generation than previously (i.e., 8 hours earlier) and 22% false alarm rate based on the analysis of historical pressure and flow data from 5 DMAs for a 3-month trial period. Therefore, based on the results obtained here, it can be stated that by using the ERS, performance improvements in terms of false alarms rate and especially

detection time (typically 15 to 30 minutes) over previously developed methods can be obtained. However, it has to be stressed that direct comparison of the results obtained in the aforementioned approaches is not possible as the tests carried out in these studies involved different DMAs and events. Furthermore, it can be noted that tests on live data/events (e.g., Mounce et al. 2010; Mounce and Boxall 2010) are more challenging than tests performed by using historical data/events.

Discussion

In the data pre-processing module of the *Setup* subsystem, outliers that are not related to bursts or other events and that should really be part of the 'NOP data sets' may be eliminated, could render the ERS detection framework more sensitive to the presence of such outliers (hereafter referred to as 'normal' outliers) in the observed data streams. This is because those pre-processed data series are then employed to train the ANN models used in the *DBA* subsystem and to compute the relevant statistics used in the *BBA* subsystem. Having said this, it can be noted that the 'normal' outliers usually occur at random times and affect a single data point in time, as opposed to burst and other events induced outliers that usually affect multiple, consecutive data points in time (and may exhibit an identifiable or quantifiable pattern). Bearing this in mind, in the ERS, the potential false positive alarms resulting from single outliers are reduced by using SPC Control Rules to analyse the incoming data over consecutive time steps. Indeed, by using these rules, persistent outliers have a greater chance of triggering an alarm than single outliers. Additionally, in the ERS, the potential false positive alarms resulting from single outliers are further reduced by using the DMA level BIS, via event occurrence evidence processing and the use of a threshold on estimated event occurrence probabilities.

In addition to the above, it could also be argued that the persistent outliers themselves may not be related to events that are of interest for water companies (e.g., bursts, sensor faults, illegal consumptions, etc.) and that should really be part of the NOP data set. These may be due to increased/decreased consumption during a very hot day, an holyday or a special occasion such as a big sport event in town, for example. At the present, the ERS does not account for these situations (although this would be possible and could be quite straightforward - e.g., information about holydays might be considered when assembling the NOP data sets and/or as an extra input to the ANN prediction model and/or as an extra input to the Bayesian Inference Systems - BISs). As a consequence, such outliers might trigger detection alarms. Bering this in mind, it is worth mentioning that the aforementioned situations would likely result in a number of alarms being simultaneously raised in multiple DMAs that cannot however be affected by the same local anomaly (e.g., a burst). Therefore, following aposteriory analysis of the raised alarms (e.g., by means of an additional ERS module or manually by the operator) it could be possible to attribute the cause of the alarms to some global anomaly (e.g., a very hot day).

CONCLUSIONS

The pressure and flow measurements collected by modern SCADA systems provide a potentially useful source of information for detecting pipe bursts and other similar events in WDSs. A novel data analysis methodology that makes use of these measurements has been presented in this paper.

The presented methodology is implemented in an automated ERS which enables the detection of pipe bursts and other events at the DMA level and as they occur. This is done by effectively and efficiently analysing the multiple and different DMA signals in a synergistic manner using a range of different AI and statistical techniques. These techniques are data-driven and have the self-learning capability. Hence, they can be dynamically recalibrated as conditions in the WDS change.

The novel event detection methodology provides the following output: (i) a probability of an event occurring in the DMA, (ii) specific probabilities for each DMA signal analysed and for each type of analysis performed, and (iii) estimates of the pressure and/or flow magnitude variations, including the likely burst flows. Furthermore, assuming that the events are quickly detected, it also provides an indication of their timing. The information generated by the novel methodology presented here may enable water companies to react more quickly after the events occurrence, mitigate their negative impacts, further diagnose the events occurring (i.e., determine their likely cause), rank alarms, and prioritise responses.

The novel event detection methodology has been tested, verified and demonstrated on a case study involving several DMAs in the UK, on both simulated burst events and real-life bursts/other events. The results obtained have shown that the ERS can detect these events effectively, reliably, and quickly thus demonstrating its potential to advance the state-of-the-art in near real-time WDSs incident management.

Note that the minimum size of the events detectable by the presented methodology depends on the DMA size, i.e., the larger the size of the DMA the larger an event has to be in order to be detected. This is because in larger DMAs the “legitimate” variability in pressure/flow may overwhelm the changes in pressure/flow due to an event. Furthermore, the pressure signals may not be affected by an event if the pipe network is large and the event is small and far away from the sensors. This said, it could be noted that with more (than one) pressure sensors within a DMA, the accuracy could improve and the DMA size may not be an issue. Also, the aforementioned “legitimate” variability in pressure/flow is less accentuated during the night hours and, it could be noted that events such as pipe burst often occur at night when the pressure in the WDS is at its highest.

Finally, note that the novel event detection methodology has been tested and verified, so far, on the UK case only and its application to pipe networks in other countries would require further tests. Also

the values of some of the event detection methodology's parameters used for the case study analyses performed here have been determined experimentally, based on a series of sensitivity analysis. This can be further improved in the future by using past recorded and verified events to semi-automatically and periodically (re)calibrate these parameters, thus leading to a more reliable detection of the DMA bursts and other events.

ACKNOWLEDGEMENTS

This work is part of the first author's PhD sponsored by the University of Exeter. The data used in the paper have been collected as part of the Neptune project funded by the UK Engineering and Physical Sciences Research Council (EP/E003192/1) and provided by Mr Ridwan Patel from Yorkshire Water which is gratefully acknowledged. The work presented in this paper has been patented (Publication No. WO/2010/131001).

REFERENCES

Adamowski, J. F. (2008). "Peak daily water demand forecast modeling using artificial neural networks." *Journal of Water Resource Planning and Management*, 134(2), 119-128.

Aksela, K., Aksela, M., and Vahala, R. (2009). "Leakage detection in a real distribution network using a SOM." *Urban Water Journal*, 6(4), 279-289.

Bates, B. C., Kundzewicz, Z. W., Wu, S., and Palutikof, J. P. (2008). *Climate change and water*. Technical Paper of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva.

Bishop, C. M. (1995). *Neural networks for pattern recognition*, Oxford University Press.

Breyfogle, F. W. (1999). *Implementing six sigma: smarter solutions using statistical methods*, Wiley, New York.

Donoho, D. L., and Johnstone, I. M. (1995). "Adapting to unknown smoothness via wavelet shrinkage." *Journal of the American Statistical Association*, 90(432), 1200-1224.

Edwards, D. (2000). *Introduction to graphical modelling, 2nd edition*, Springer.

Egan, J. P. (1975). *Signal detection theory and ROC analysis*, Series in Cognition and Perception, Academic Press, New York.

Fenner, R. A., and Ye, G. (2011). "Kalman filtering of hydraulic measurements for burst detection in water distribution systems." *Journal of Pipeline Systems Engineering and Practice*, 2(1), 14-22.

Grunwell, D., and Ratcliffe, B. (1981). *Location of underground leaks using the leak noise correlator*, Water Research Centre, Technical Report 157.

Hanley, J. A., and McNeil, B. J. (1982). "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology,* 143, 29-36.

Kapelan, Z., Savić, D. A., and Walters, G. (2003). "A hybrid inverse transient model for leakage detection and roughness calibration in pipe networks." *Journal of Hydraulic Research,* 41(5), 481-492.

Mallat, S. (1989). "A theory for multiresolution signal decomposition: the wavelet representation." *IEEE Pattern Analysis and Machine Intelligence*, 11(7), 674-693.

Mergelas, B., and Henrich, G. (2005). "Leak locating method for precommissioned transmission pipelines: North American case studies." *Proc. Leakage 2005 conference*, Halifax, Canada.

Meyer, Y., and Roques, S. (1993). *Progress in wavelet analysis and applications*, Frontières editions.

Misiunas, D., Vítkovský, J., Olsson, G., Simpson, A., and Lambert, M. (2005). "Pipeline burst detection and location using a continuous monitoring of transients." *Journal of Water Resources Planning and Management*, 131(4), 316-325.

Mounce, S. R., and Boxall, J. B. (2010). "Implementation of an on-line artificial intelligence district meter area flow meter data analysis system for abnormality detection: a case study." *Water Science and Technology: Water Supply,* 10(3), 437-444.

Mounce, S. R., Boxall, J. B., and Machell, J. (2010). "Development and verification of an online artificial intelligence system for burst detection in water distribution systems." *Journal of Water Resources Planning and Management*, 136(3), 309-318.

Mounce, S. R., Mounce, R. B., and Boxall, J. B. (2011). "Novelty detection for time series data analysis in water distribution systems using support vector machines." *Journal of Hydroinformatics*, 13(4), 672-686

Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models, Part I - A discussion of principles." *Journal of Hydrology*, 10(3), 282-290.

Neuroshell2 manual (1996). Ward Systems Group Inc.

Palau, C. V., Arregui, F. J., and Carlos, M. (2011). (Accepted) "Burst detection in water networks using principal component analysis." *Journal of Water Resources Planning and Management*.

Puust, R., Kapelan, Z., Savić, D. A., and Koppel, T. (2010). "A review of methods for leakage management in pipe networks." *Urban Water Journal*, 7(1), 25-45.

Romano, M., Kapelan, Z., and Savić, D. A. (2010a). "Pressure signal de-noising for improved real-time leak detection." *Proc. 9th International conference on Hydroinformatics 2010*, Tianjin, China.

Romano, M., Kapelan, Z., and Savić, D. A. (2010b). "Real-time leak detection in water distribution systems." *Proc. 12th Water Distribution Systems Analysis conference*, Tucson, Arizona.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). "Learning internal representations by error propagation." *Parallel distributed processing: explorations in the microstructure of cognition, vol. 1: foundations*, MIT Press, 318-362.

Shewhart, W. (1931). *Economic control of quality of manufactured product*, Van Nostrand Reinhold.

Western Electric Company (1958). *Statistical quality control handbook, 2nd edition*, AT&T Technologies.

LIST OF FIGURES

LIST OF TABLES

1 Table 1. Parameters used in the case study.

Schedule & data retrieval	l=7 days; m=14 days; a=30minutes; b=15values; c=11values; d=4 hours; e=14 days; f=90 days; g=96; v_{ph}=4.
Statistical tests	$N_{l1^{st}} = N_{u1^{st}} = N_{l2^{nd}} = N_{u2^{nd}} = 3.$ $N_{l3^{rd}} = N_{u3^{rd}} = N_{l4^{th}} = N_{u4^{th}} = 3.$
ANN models	*LagSize*=4 values; *Train%*=80%; *Test%*=20%;
BISs	z=3; λ=0.5.

2

Table 2. Engineered Events time schedule, detection times, actual and estimated flow rates, and actual flow rate as a percentage of the average District Metered Area inflow.

	Alarm start time	Hydrant opened time	Hydrant closed time	Actual flow rate	% average inflow	Estimated flow rate
July	20/07/2009 - 08:00	20/07/2009 - 08:00	21/07/2009 - 08:00	2 l/s	11%	1.9 l/s
	22/07/2009 - 06:15	22/07/2009 - 06:05	23/07/2009 - 08:00	2 l/s	11%	2 l/s
	24/07/2009 - 16:15	24/07/2009 - 16:00	25/07/2009 - 16:00	2 l/s	11%	2 l/s
August	17/08/2009 - 08:30	17/08/2009 - 08:20	18/08/2009 - 07:05	3 l/s	16%	2.7 l/s
	21/08/2009 - 08:00	21/08/2009 - 07:20	22/08/2009 - 08:05	1 l/s	5%	0.8 l/s
	26/08/2009 - 08:00	26/08/2009 - 07:55	27/08/2009 - 07:00	1 to 3 l/s	5%-16%	2.2 l/s
March	**01/03/2010 - 21:00**	01/03/2010 - 09:10	02/03/2010 - 07:50	2 l/s	11%	2.1 l/s
	02/03/2010 - 21:00	02/03/2010 - 08:00	03/03/2010 - 07:10	2 l/s	11%	2 l/s
	15/03/2010 - 07:45	15/03/2010 - 07:20	16/03/2010 - 07:15	2 l/s	11%	2 l/s

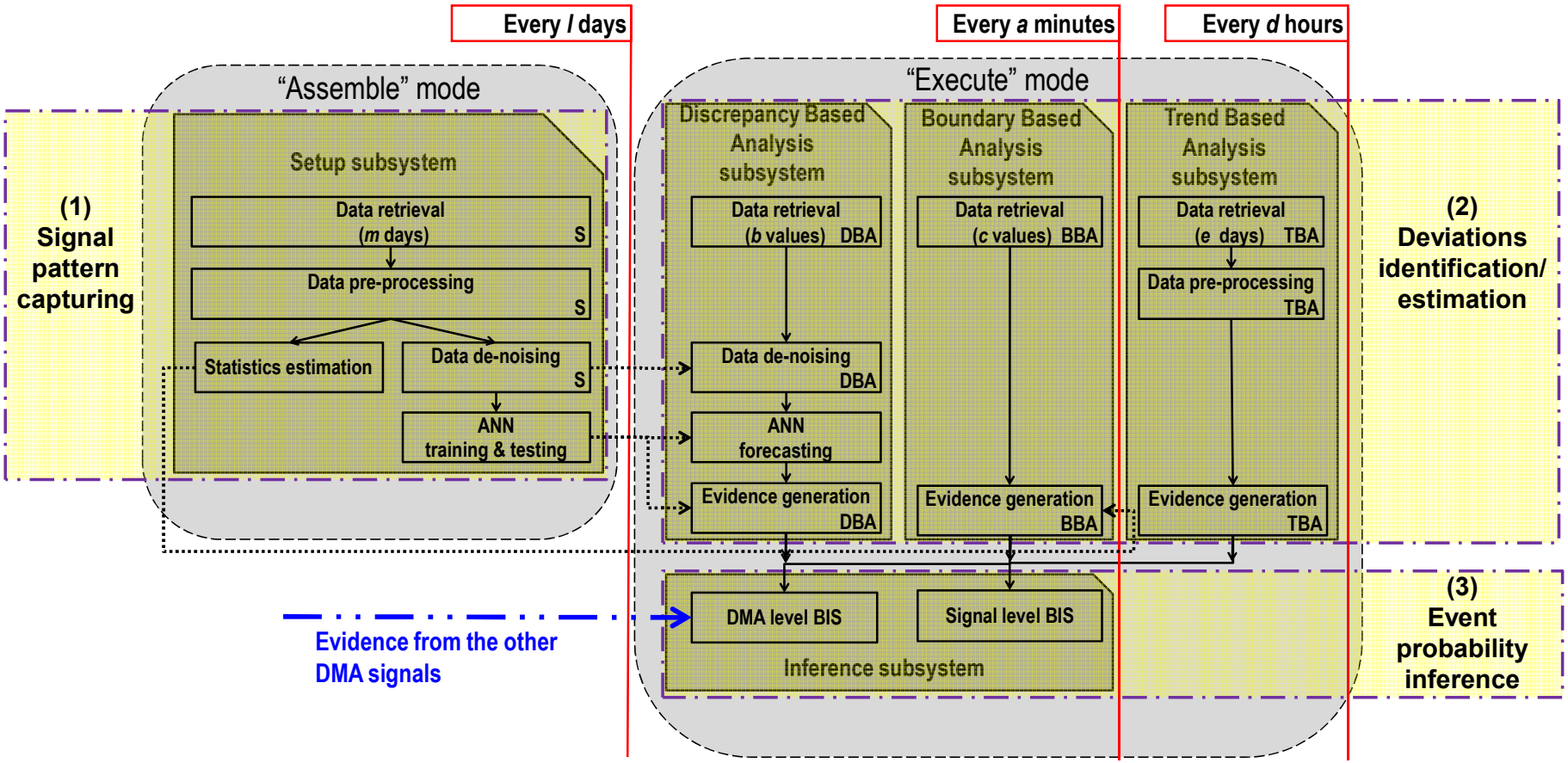

Figure 1. Diagrammatic representation of the Event Recognition System.

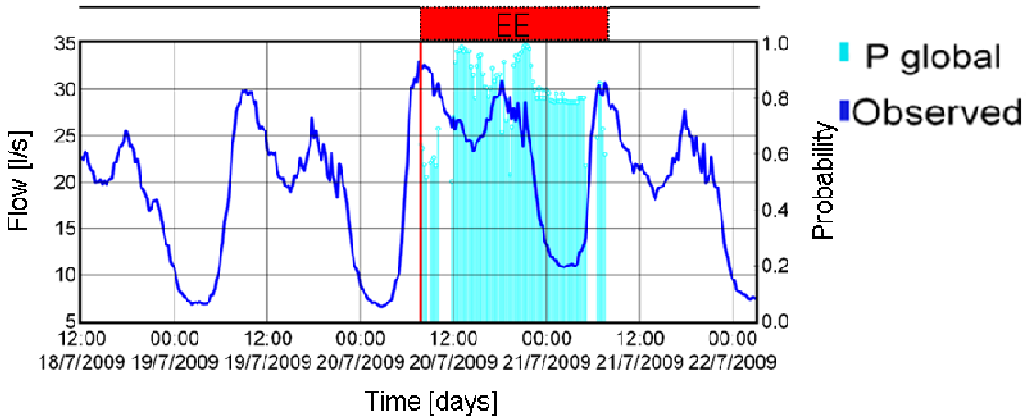

Figure 2. Results obtained by the Event Recognition System when an Engineered Event occurred.

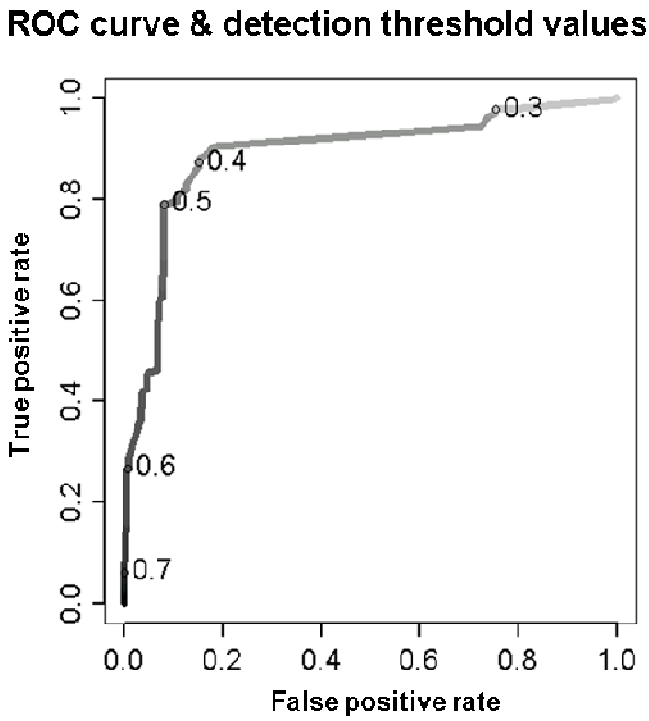

Figure 3. Receiver Operating Characteristics graph for the three month data set and detection threshold values.

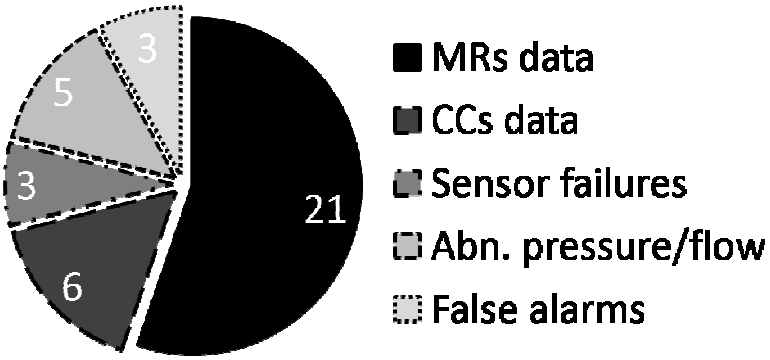

Figure 4. Number of alarms correlated with the events identified by visual inspection of the data.